The Development and Standardization of the I.J.R. Test for the Visually Handicapped

Margaret Davidson & Andrew W. Brown

BF 431
D 283

PAMPHLET BINDERS

This is No. 1525

also carried in stock in the following sizes

	HIGH		WIDE		THICKNESS
1523	9	inches	7	inches	½ inch
1524	10	"	7	"	"
1525	9	"	6	"	"
1526	9¾	"	7⅛	"	"
1527	10½	"	7⅜	"	"
1528	11	"	8	"	"

	HIGH		WIDE		THICKNESS
1529	12	inches	10	inches	½ inch
1530	12	"	9⅛	"	"
1932	13	"	10	"	"
1933	14	"	11	"	"
1934	16	"	12	"	"

Other sizes made to order.

MANUFACTURED BY

LIBRARY BUREAU

Division of REMINGTON RAND INC.

Library Supplies of all kinds

AMERICAN FOUNDATION
FOR THE BLIND INC.

BF431
D283
Cop. 1

Journal of Applied Psychology
v.XXIII No.2 Apr. 1939

THE DEVELOPMENT AND STANDARDIZATION OF THE I.J.R. TEST FOR THE VISUALLY HANDICAPPED[1]

MARGARET DAVIDSON AND ANDREW W. BROWN
Institute for Juvenile Research, Chicago

THERE are, for purposes of general clinical use, a large number of fairly satisfactory tests for the measurement of so-called "general intelligence." Most children may be given any of a large number of individual or group, language or non-language tests intended to determine their standing in this particular respect. Unfortunately, however, when the practicing psychologist finds himself confronted with a child handicapped with respect to sight or hearing, he is not so well supplied with instruments. It is the individual with a visual handicap with which we have concerned ourselves in this study.

The Hayes-Binet is the one test that has been recognized and rather generally used with subjects with visual handicaps. This test, however, is designed for use only with the totally blind, which means that it is not suitable for the largest group of visually handicapped subjects—those with enough vision so that they have not learned Braille, and who do not, in general, employ the same techniques for getting along in the world as do the totally blind.

The purpose of this study has been to develop a test by which the same thing measured by the conventional intelligence test may be measured in those who are visually handicapped; not only in those who are blind, but also in those whose vision is so defective that they might conceivably be at a disadvantage in performing on any of the usual scales.

---

[1] The test described herein may, for the present, be secured by writing to Andrew W. Brown, Institute for Juvenile Research, 907 S. Wolcott Avenue, Chicago, Illinois.

The writers wish to express their thanks to Dr. M. W. Richardson for his assistance with some of the technical problems involved in the construction of the test.

We wish also to acknowledge clerical assistance from W.P.A.

Studies from the Institute for Juvenile Research, Chicago, Paul L. Schroeder, M.D., Director. Series C, No. 187.

An attempt has been made so to select the material and procedure that the test can be used with subjects whose vision is of any degree whatever. They may be totally blind, or completely without visual handicap, so far as the suitability of the test is concerned. The norms tentatively established at this time, and those to be adopted later, are, and will be, based on the scores of a group that is normal with respect to vision.

For a number of reasons the norms of the test have been based upon groups assumed to be visually normal. *First,* it would be extremely difficult, both in standardizing and in using the test, to draw the lines of demarcation between different degrees of visual handicap for purposes of the choice of scale. *Second,* the sampling problem, difficult enough in dealing with the general population, becomes much more so in dealing with such a sub-population as that composed of the visually handicapped. In the first place the difficulty of obtaining a sufficiently large number of subjects for purposes of establishing norms without going into institutions for the visually handicapped, which is obviously undesirable, would be considerable. In the second place, it seems reasonable that any norms truly representative of a handicapped group would necessarily shift very considerably from time to time and place to place as a result of variations in treatment and methods of prevention of visual disease. *Third,* it is more useful to be able to compare the intelligence of the handicapped subject with that of the visually normal subject, as well as with that of others who are handicapped, than to be able to compare him only with others who are visually handicapped. And *finally,* if we establish a way of equating the two types of individuals with respect to general intelligence it will become possible to compare and contrast similar groups of them with respect to other abilities and traits.

We have chosen, further, to set up the test in the point scale form, wishing to avoid the difficulties and ambiguities to which the use of the mental-age scale tends to lead. As to the general method of administration of the test there has been little choice, its purpose determining this to a very large extent. It is of necessity an individual test. The instructions are given orally in accordance with carefully worked out directions, and the subject gives his answers orally, the examiner recording as much of such answers as is thought necessary.

The material used in the test is not, for the most part, original. The first selection of tests was made simply by going through all available accepted tests of intelligence and selecting from them the type of material which seemed on inspection to be of the sort desired; that is, material was selected which would probably be equally difficult for children of all degrees of vision. This procedure yielded a large quantity of material from which to choose. This was then sifted for tests of such a type as could be developed into point scales with widely varying degrees of difficulty. The tests thus selected are:

I. Repetition of sentences.
II. Repetition of digits in reversed order.
III. Vocabulary, concrete.
IV. Same-opposites.
V. Absurdities.
VI. Opposites.
VII. Differences.
VIII. Word naming.
IX. Repetition of the thought of a passage.
X. Similarities.
XI. Disarranged sentences.
XII. Arithmetic problems.
XIII. Logical inferences.
XIV. Analogies.
XV. Number series completion.
XVI. Vocabulary, abstract.
XVII. Comprehension of passages read to subject.

These tests are all, except for test XVII, conventional and frequently used tests of intelligence. They were used here as nearly as possible in their usual forms, although in many cases it was necessary to alter procedures and methods of administration in order to adapt them to a completely oral mode of administration. Test XVII involves the reading to the subject by the examiner of three paragraphs, after each of which the subject is asked to answer five or six questions.

Specific items of each of the tests were so selected that they might be, as nearly as could be determined by inspection, equally difficult for the blind and the seeing. Various persons experienced in dealing with the blind, and several blind people were consulted during this process, and in the selection of material their suggestions were followed wherever possible. It was necessary, however, to have some further evidence that the material was equally well adapted to the several types of subjects. The seventeen tests were therefore given to three groups of subjects: 65 partially blind pupils in the sight-saving classes of the Chicago public schools; 45 pupils in the classes for the blind in the Chicago public schools; and 21 pupils with normal vision in the Illinois Soldiers' and Sailors' Children's School. Other ratings were obtained for all members of the three groups; the partially blind were given the Stanford-Binet examination—tests being omitted or alternates used at the discretion of the examiners in such a manner as to obtain what seemed to be the best possible rating; the blind were given both the Stanford and Hayes-Binet tests; and Stanford-Binet ratings were available for the seeing subjects of the Illinois Soldiers' and Sailors' Children's School. The blind and seeing groups were matched, the average I.Q. and spread of I.Q.'s being as nearly as possible equated within each age group.

Comparison of the scores of the three groups on the tests seemed to indicate that we were justified in assuming that there were no significant differences in the ability of the blind, partially blind, and seeing to perform tasks of the type represented by the seventeen tests. This conclusion was based on two things: (1) a comparison of the means for the seeing and blind groups on each test and computation of the actual difference between means with the probable error of the difference; (2) the observation of the number of age levels at which either group showed a higher average score on each test. In general there was a slight but insignificant tendency for the blind group to obtain higher scores than the seeing.

On the basis of this analysis the data for all three groups were pooled and each of the 17 subtests was studied and compared with the others in order that a smaller number of tests might be selected to make up the final battery. The correlation of each test with Stanford-Binet mental age was calculated, inter-test correlations were obtained, and scoring methods were studied. Correlation with Stanford-Binet mental age was considered most important in the final selection of the tests, but other things taken into consideration were: objectivity of scoring, length of time required to administer, and apparent equality of difficulty for the three groups. On these bases ten tests were retained:

1. Vocabulary (abstract and concrete combined).
2. Comprehension.
3. Arithmetic problems.
4. Repetition of the thought of a passage.
5. Repetition of digits reversed.
6. Opposites.
7. Similarities.
8. Disarranged sentences.
9. Number sequence.
10. Analogies.

These subtests were considerably revised, and two forms of each were developed. The *vocabulary* test incorporated into the final form is completely different from that in the original form. This revision was made quite largely in an effort to use words more easy to score. The *arithmetic problems* test was cut into two equivalent halves, as were the *opposites, similarities,* and *analogies* tests. Parallel forms of the *comprehension, repetition of the thought of a passage, digits, disarranged sentences,* and *number sequence* tests were devised.

Each form of the *vocabulary* test consists of 20 items ranging in difficulty from such words as *table* to such words as *macrocosm.* One point of credit is given for each correct response. Differential scoring was tried for a time, but it was found to be so unreliable that it was dropped.

Each form of the *comprehension* test contains 16 questions based on three paragraphs of widely varying difficulty. The first paragraph consists, in each case, of very simple narrative material; the second of expository material; and the third of abstract material of considerable difficulty. One point of credit is given for each question answered correctly, and in several cases partial credit is possible. The scoring of the test is fairly easy.

Each form of test 3 contains ten arithmetic problems varying in difficulty from such questions as "If you have two pennies and somebody gives you another, how many pennies do you have all together?" to "If a wire twenty inches long is to be cut so that one piece is two-thirds as long as the other piece, how long must the longer piece be?" One point of credit is given for each correct response.

Each form of the *repetitions of the thought of a passage* test consists of five paragraphs which are read to the subject, and after each he is asked to repeat in his own words as much of the paragraph as he can. The paragraphs vary from very simple descriptive material to such as that contained in the two paragraphs in the old Stanford-Binet examination. For scoring purposes each one of the paragraphs is broken up into small divisions, each of which is supposed to contain a single idea. The subject is given one point of credit for each of these divisions the idea of which has been contained in his response. The scoring of this test is difficult as it is quite subjective in nature. However, detailed instructions for scoring have been worked out with examples, and if these are carefully studied and followed the test will be decidedly useful, even though not perfectly reliable with respect to scoring.

The test of *repeating digits reversed* is the conventional test of this type. There are eight items in each form, the shortest containing combinations of two digits, the longest, nine. One point of credit is given for each item passed.

Each form of the *opposites* test contains ten items, varying in difficulty from *up* to *pride.* One point is given for each opposite correctly named.

Each form of the *similarities* test contains ten items varying in difficulty from such combinations as *apple-peach* to such as *deafness-feeblemindedness-deceitfulness.* One point of credit is given for each correct response.

Each form of the *disarranged sentences* test contains eleven sentences varying in difficulty from such combinations as "dog the barks" to "certain always death of cause kinds sickness." Partial credit is given on this test in accordance with definite rules. Scores of 0, 1, and 2 are possible on each sentence.

Each form of the *number sequence* test contains ten items varying in difficulty from such items as 5, 10, 15, 20, —, — to such as 3, 4, 6, 10, —, —. The subject is asked to give the next two numbers in the sequence. For each number which is correct and in its correct position one point of credit is given. Thus the highest possible score on this test is 20.

One form of the *analogies* test contains 11 items, the other 12. These items vary in difficulty from such analogies as Monday : Tuesday : : Friday : ———————— to; point : line : : line : ————————. One point of credit is given for each correct response.

Detailed instructions for the giving and scoring of the test have been worked out. The average time required for giving each form is probably about an hour, although, as is true of most such tests, the younger the subject, the shorter the time required. With less mature subjects it is not necessary to give all of each subtest. The limits of testing for each test have been worked out on the basis of the data on hand.

The total score is the weighted sum of the ten subtest scores, the following weights being given to each:

Sub-Test	Wt.	Sub-Test	Wt.
1	3	6	4
2	2	7	4
3	4	8	2
4	1	9	2
5	5	10	3

These weights, and most of the remainder of our present description of the test, are based on the test scores of 418 subjects with normal vision. These subjects were for the most part pupils in the Chicago public schools. The age distribution of the group follows:

Age	Frequency	Age	Frequency
6–0 to 6–5	9	13–6 to 14–5	33
6–6 to 7–5	22	14–6 to 15–5	35
7–6 to 8–5	13	15–6 to 16–5	23
8–6 to 9–5	27	16–6 to 17–5	12
9–6 to 10–5	37	17–6 to 18–5	12
10–6 to 11–5	62	18–6 to 19–5	4
11–6 to 12–5	66	19–6 and above	8
12–6 to 13–5	55		

The grade distribution is as follows:

Grade	Frequency	Grade	Frequency
1	21	7	52
2	23	8	56
3	19	9	49
4	25	10	12
5	63	11	9
6	69	12	10

Of the ten subjects not included in the grade frequency distribution, six were not at the time in school. Of these, two reported having completed three years of high school, three reported having completed high school, and one had completed one year in college. The educational history of the remaining four is unknown.

The following table (p. 237) gives the number of subjects at each age level who reported the use of a foreign language in the home, and the number of Negro subjects at each age level. The only bases for the selection of these cases were that they were drawn from schools in neighborhoods in which the popu-

Age	Number reporting use of language other than English	No. of Negro subjects
6–0 to 6–11	3	
7–0 to 7–11	6	
8–0 to 8–11	4	
9–0 to 9–11	6	1
10–0 to 10–11	11	23
11–0 to 11–11	14	31
12–0 to 12–11	17	14
13–0 to 13–11	14	9
14–0 to 14–11	2	7
15–0 to 15–11	5	8
16–0 to 16–11		
17–0 to 17–11		5
18–0 to 18–11		3
19–0 and above		2

lation was of approximately average economic status and that, in general, children whose grade placements deviated markedly from that indicated by their chronological ages were avoided. It is recognized that the group is not ideal, particularly in that there is such a disproportionally large number of Negro subjects at some age levels. The norms based on the group are, of course, only tentative.[2]

The weights listed for the ten subtests are based on the inter-test correlations for this group and are roughly proportional to the quantity

$$W_i = \frac{\sum_{i=1}^{10} r_{ij}}{\sigma_i}$$

[2] For the revision and extension of the norms ten schools have been selected which as far as could be determined from the census report give a random sampling of the population. Every seventh child on the teacher's register in each grade in each school has been examined. Thirteen hundred and seven subjects have been given both forms of the test. The norms will be based on the age distribution of these cases and the 418 cases included in this report.

where the numerator represents the sum of the correlations of test i with all ten tests, its self-correlation being estimated as one, and the denominator represents the standard deviation of scores for test i. The use of this formula has the effect of weighting the subtests in proportion to the extent to which they correlate with that which is measured by the test as a whole. This formula, and that for the test reliability, which follows, were developed by Dr. M. W. Richardson in work as yet unpublished.

The reliability of the test was estimated from the calculated correlations between forms A and B of each of the ten subtests according to the formula

$$1 - R_{tt} = \frac{\sum_{i=1}^{10} w_i^2(1 - r_{ii})}{\sum_{i=1}^{10} w_i^2 + 2\sum_i \sum_j w_i w_j r_{ij}} \quad i \neq j$$

a formula for estimating the reliability coefficient of the weighted sum of the several subtests. In this formula R_{tt} is the reliability estimate, the w_i's are the products of the weights for each test i by the standard deviations for the corresponding test i, the r_{ii}'s are the correlations between forms A and B of each test i, and the r_{ij}'s are the correlations of each test i with every other test j. The reliability coefficient for the total test as estimated by this formula is .96. The Pearson product-moment correlation between scores on form A and form B of the total test is .95. It seems fairly certain, then, that the reliability of the test is satisfactory.

The correlation of scores on form A with chronological age is .72; that of form B with chronological age is .70. These correlations are based on the 341 subjects ranging in age from 6–0 through 14–11.

The standard deviation of the test scores for the total group of 418 cases is 59 points. The contribution made by each of the ten subtests to the total variance of the test is as follows:[3]

Sub-test	% variance contributed	Sub-test	% variance contributed
1	10.1	6	11.8
2	9.4	7	8.8
3	9.2	8	13.2
4	11.1	9	9.8
5	7.3	10	9.3

[3] These figures are based on the formula

$$\sigma_t{}^2 = w_1 \sum_{i=1}^{i=10} w_i\, r_{1i} + w_2 \sum_{i} w_i\, r_{12} + \;.\;.\;.\;.\;.\; + w_{10} \sum_{i} w_i\, r_{10i}$$

where each term in the right member represents the variance contribution of the subtest with corresponding subscript. This formula is taken from unpublished material by Dr. M. W. Richardson and was secured from him personally.

EMPLOYEE SELECTION TESTS FOR ELECTRICAL FIXTURE ASSEMBLERS AND RADIO ASSEMBLERS*

JOSEPH TIFFIN AND R. J. GREENLY
Purdue University

IN initiating a program of psychological testing in plants which have not previously used tests of this type in their employment departments, it has seemed advisable in each case to "test the tests" on present operators before making a definite recommendation as to which tests are suitable for specific jobs. The advantages of this procedure over recommending tests on the basis of a logical analysis of the job are several. Tests which do not function are practically certain to be eliminated. In the present stage of psychological testing, a battery of tests selected without such experimentation is almost certain to contain, at least occasionally, one or more tests which do not function. One such mistake will harm a program in the eyes of the management enough to counteract the favorable impression created by many correct choices. Secondly, it is very difficult for a psychologist who has not *lived* with the job to obtain by observation a knowledge of the operations intimate enough to enable him to select from a number of possible tests the three or four which are most suited to that particular job. Finally, a plant manager or superintendent usually insists on seeing the proof of any new idea before he installs it. If he can be shown that a competent psychologist can come into his plant with a battery of tests and select those workers who are known from production records or ratings to be most efficient, he will usually agree that the

---

* Because of the growing interest of industrial executives and operating supervisors in psychological methods of personnel selection, this article is addressed primarily to them and deals at greater length with certain concepts and statistical procedures than is customary in a professional journal. *Editor.*

c. 1

BF431 DAVIDSON, MARGARET, &
D283 BROWN, ANDREW W. The development and standardization of the....

Date Due

---

c. 1

BF431
D283

AUTHOR DAVIDSON, MARGARED, &
BROWN, ANDREW W.

TITLE THE DEVELOPMENT AND STANDARDIZATION OF THE I.J.R.TEST FOR THE VISUALLY HANDICAPPED.

DATE DUE	BORROWER'S NAME
	Reference Copy

VOC. REHAB DIV.
AUSTIN, TEXAS